POPULAR DIFFERENCE SETS

TOM SANDERS

Abstract. We provide further explanation of the significance of an example in a recent paper of Wolf in the context of the problem of finding large subspaces in sumsets.

Throughout these notes G will denote the group $\mathbb{F}_2^n$, that is the n-dimensional vector space over $\mathbb{F}_2$, and $\mathbb{P}_G$ will denote the normalized counting measure on G. The convolution of two functions $f, g \in L^1(\mathbb{P}_G)$ is defined to be

$$f * g(x) := \int f(y)g(x-y)d\,\mathbb{P}_G(y),$$

and is of particular importance in additive combinatorics. To see why recall that if $A, B \subset G$ then we write $A+B$ for the sumset $\{a+b : a \in A, b \in B\}$. Convolution then yields the following important identity

$$A + B = \operatorname{supp} 1_A * 1_B.$$

The sumset $A+B$ can thus be studied through the convolution $1_A * 1_B$; a typical example of this is in the following problem.

Problem 1.1 (Bourgain-Green, [Bou90, Gre02, Gre05])**.** Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\Omega(1)$. How large a subspace V with $V \subset A + A$ can we guarantee may be found?

All known approaches to this problem proceed by harmonic analysis of $1_A * 1_A$. Since the arguments are analytic the methods are not good at distinguishing between when $1_A * 1_A$ is small and when it is very small. To be more precise we introduce a definition.

Given a set $A \subset G$ of density $\alpha > 0$ and a parameter $c \in [0,1]$ we define the *popular difference set* with parameter c to be

$$D_c(A) := \{x \in G : 1_A * 1_A(x) > c\alpha^2\}.$$

Our definition employs a slightly different normalization to those presented elsewhere so that c naturally lies in the range $[0,1]$. Indeed, it is easy to see that if c is greater than 1 then even for large sets A we may have $D_c(A) = \{0_G\}$.[1]

At the other end of the spectrum we have $D_0(A) = A + A$. Now, the arguments for tackling Problem 1.1 all fail to meaningfully distinguishing between the set $A + A = D_0(A)$ and $D_c(A)$ when c is small. In particular, for example, Green effectively proves the following result in [Gre05].

[1] In particular this occurs generically: if $c > 1$, $|G|$ is large enough in terms of c, and A is chosen uniformly at random from sets of size $|G|/2$ then (with high probability) $D_c(A) = \{0_G\}$.

Theorem 1.2. *Suppose that $G := \mathbb{F}_2^n$, $A \subset G$ has density $\alpha = \Omega(1)$ and $c \in [0,1]$ is a parameter. Then there is a subspace $V \subset D_c(A)$ with*

$$|V| \geq \exp(\Omega((1-c)n)).$$

Notice that the bound necessarily tends to 1 as c tends to 1, but when c is small there is no significant variation. Setting $c = 0$ one has the following corollary and nothing stronger is known.

Corollary 1.3 ([Gre05, Theorem 9.3])**.** *Suppose that $G := \mathbb{F}_2^n$ and $A \subset G$ has density $\alpha = \Omega(1)$. Then there is a subspace $V \subset A + A$ with*

$$|V| \geq \exp(\Omega(n)).$$

In the other direction, adapting a construction from Ruzsa [Ruz91] (see also [Ruz87]), Green showed the following result.

Theorem 1.4 ([Gre05, Theorem 9.4])**.** *Suppose that $G := \mathbb{F}_2^n$. Then there is a set $A \subset G$ with density $\alpha = \Omega(1)$ such that if $V \subset A + A$ is a subspace then*

$$\mathbb{P}_G(V) \leq \exp(-\Omega(\sqrt{n})).$$

Of course the gap between Corollary 1.3 and Theorem 1.4 is very large so in a sense Problem 1.1 remains wide open.

Since our only ways of proving that $A + A$ contains a subspace also show that $D_c(A)$ contains a subspace it is natural to ask about the limitations of such methods and, in particular, ask Problem 1.1 with $D_c(A)$ in place of $A + A$.

It turns out that $D_c(A)$ need not contain a large subspace. Indeed, more than this, in the paper [Wol09] Wolf was able to show that there are sets where not only does $D_c(A)$ not contain a large subspace, it doesn't even contain the sumset of a large set.[2]

Theorem 1.5 ([Wol09, Theorem 2.6])**.** *Suppose that $G := \mathbb{F}_2^n$ and $c \in (0, 1/2]$ is a parameter. Then there is a set $A = A(c) \subset G$ with density $\alpha = \Omega(1)$ such that if $A' + A' \subset D_c(A)$ then*

$$\mathbb{P}_G(A') \leq \exp(-\Omega(n/\log^2 c^{-1})).$$

We remind the reader that we are interested in the case when c is small but fixed; as c tends to 0 the upper bound must tend to 1 since $A + A = D_0(A)$.

The proof makes appealing use of measure concentration in G. Indeed, it is a key insight here to consider the more general question of containing the sumset of a large set, not just a subspace, as this is much more suggestive of such tools.

Problem 1.6 (Wolf, [Wol09])**.** Suppose that $A \subset G := \mathbb{F}_2^n$ has density $\Omega(1)$ and $c \in [0,1]$ is a parameter. How large a set A' with $A' + A' \subset D_c(A)$ can we guarantee may be found?

Remarkably it turns out that it is quite easy to see that Wolf's theorem is close to best possible; we have the following complementary result.

Theorem 1.7. *Suppose that $G := \mathbb{F}_2^n$, $A \subset G$ has density $\alpha = \Omega(1)$ and $c \in (0,1]$ is a parameter. Then there is a set A' such that $A' + A' \subset D_c(A)$ and*

$$\mathbb{P}_G(A') \geq \exp(-O(n/\log c^{-1})).$$

---

[2]In actual fact the bulk of [Wol09] concerns the arithmetic case $\mathbb{Z}/N\mathbb{Z}$; our purpose is expository and so our interest remains with $\mathbb{F}_2^n$.

This result does not quite achieve the bounds of Theorem 1.5 and it seems of interest to try to close this gap.

Since we are looking for the considerably weaker structure of a sumset rather than a subspace we have some rather stronger tools available to us in the form of Gowers' [Gow98] proof of the Balog-Szemeredi theorem [BS94]. We shall prove the following explicit version of Theorem 1.7

Theorem 1.8. *Suppose that $G := \mathbb{F}_2^n$, $A \subset G$ has density $\alpha > 0$ and $c \in (0, 1/2]$ is a parameter. Then there is a set A' such that $A' + A' \subset D_c(A)$ and*

$$|A'| \geq \lfloor \alpha^3 2^{n(1-\log \alpha^{-1}/\log c^{-1})}/12 \rfloor.$$

Note that, as expected, Gowers' method provides rather good density dependence; using Fourier methods one might expect this to be exponential in $\alpha^{-O(1)}$ rather than in $\log \alpha^{-1}$.

The following result is an obvious generalization of the 'combinatorial lemma' of Gowers [Gow98, Lemma 11].

Lemma 1.9. *Suppose that $G := \mathbb{F}_2^n$, $A \subset G$ has density $\alpha > 0$ and $c, \sigma \in (0, 1]$ are parameters. Then there is a set $A' \subset G$ with*

$$\mathbb{P}_G(A') \geq \alpha^{\lceil \log 2\sigma^{-1}/\log c^{-1} \rceil}/\sqrt{2},$$

and

$$\mathbb{P}_G^2(\{(x, y) \in A'^2 : x + y \in D_c(A)\}) \geq (1 - \sigma)\,\mathbb{P}_G^2(A'^2).$$

Proof. Let r be a natural parameter to be specified later and $X_1, \ldots, X_r$ be independent, uniform, G-valued random variables. Put $A_i = X_i + A$ and $A' = \bigcap_{i=1}^r A_i$.

The probability that $(x, y) \in A_i^2$ is $1_A * 1_A(x + y)$, so the probability that $(x, y) \in A'^2$ is $1_A * 1_A(x + y)^r$. However, $\mathbb{E}_{x,y \in G} 1_A * 1_A(x + y) = \alpha^2$ whence, by Hölder's inequality, we have that $\mathbb{E}\,\mathbb{P}_G(A')^2 = \mathbb{E}_{x,y \in G} 1_A * 1_A(x + y)^r \geq \alpha^{2r}$.

Let S be the set of pairs $(x, y) \in A'^2$ such that $x + y \notin D_c(A)$. Then, as before, the probability that $(x, y) \in S$ is $1_A * 1_A(x + y)^r \leq c^r \alpha^{2r}$ and it follows that the expectation of $\mathbb{P}_G^2(A'^2) - \sigma^{-1}\,\mathbb{P}_G^2(S)$ is at least $\alpha^{2r}(1 - \sigma^{-1}c^r)$. Letting $r = \lceil \log 2\sigma^{-1}/\log c^{-1} \rceil$ we conclude that there are some values of $X_1, \ldots, X_r$ such that $\mathbb{P}_G^2(A'^2) - \sigma^{-1}\,\mathbb{P}_G^2(S) \geq \alpha^{2r}/2$. In particular

$$\mathbb{P}_G^2(A'^2) - \sigma^{-1}\,\mathbb{P}_G^2(S) \geq 0 \text{ and } \mathbb{P}_G^2(A'^2) \geq \alpha^{2r}/2,$$

and the result follows. □

We may now prove the theorem using the pigeon-hole principle.

Proof of Theorem 1.8. Let σ be a parameter such that

$$\left\lceil |G| \alpha^{\lceil \log 2\sigma^{-1}/\log c^{-1} \rceil}/\sqrt{2} \right\rceil \geq \sigma^{-1}. \tag{1.1}$$

Apply Lemma 1.9 with σ to get a set $A_0 \subset G$ with $|A_0| \geq \sigma^{-1}$ and

$$\mathbb{P}_G^2(\{(x, y) \in A_0^2 : x + y \in D_c(A)\}) \geq (1 - \sigma)\,\mathbb{P}_G^2(A_0^2). \tag{1.2}$$

Let $A_1 \subset A_0$ be chosen uniformly at random from sets of size $\lfloor \sigma^{-1}/4 \rfloor$ (possible since $|A_0|$ is large enough). It follows that $\mathbb{E}\,|\{(x, y) \in A_1 \times A_1 : x + y \in D_c(A)\}|$ is equal to

$$\mathbb{E}\,|\{(x, y) \in A_1 \times A_1 : x \neq y, x + y \in D_c(A)\}| + |A_1|.$$

The first term here is at least

$$\frac{|A_1|(|A_1|-1)}{|A_0|(|A_0|-1)}((1-\sigma)|A_0|^2-|A_0|),$$

by (1.2). Whence

$$\mathbb{E}\,|\{(x,y)\in A_1\times A_1 : x+y\in D_c(A)\}|\geq (1-2\sigma)|A_1|^2,$$

since $|A_0|^{-1}\leq\sigma$. We may thus pick a set A_1 such that this inequality is satisfied.

Now, let

$$A_2:=\{x\in A_1 : |\{y\in A_1 : x+y\in D_c(A)\}|\geq(1-3\sigma)|A_1|\}.$$

Note that since $3\sigma|A_1|<1$, $x\in A_2$ implies that $x+A_1\subset D_c(A)$. However, $A_2\subset A_1$ whence $A_2+A_2\subset D_c(A)$. It remains to note that

$$|A_2||A_1|+(|A_1|-|A_2|)(1-3\sigma)|A_1|\geq(1-\sigma)|A_1|^2,$$

whence $|A_2|\geq|A_1|/2\geq\lfloor\sigma^{-1}/4\rfloor/2$ and we have the conclusion on maximizing this with σ subject to (1.1). □

References

[Bou90] J. Bourgain. On arithmetic progressions in sums of sets of integers. In *A tribute to Paul Erdős*, pages 105–109. Cambridge Univ. Press, Cambridge, 1990.

[BS94] A. Balog and E. Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[Gow98] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[Gre02] B. J. Green. Arithmetic progressions in sumsets. *Geom. Funct. Anal.*, 12(3):584–597, 2002.

[Gre05] B. J. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.

[Ruz87] I. Z. Ruzsa. Essential components. *Proc. London Math. Soc. (3)*, 54(1):38–56, 1987.

[Ruz91] I. Z. Ruzsa. Arithmetic progressions in sumsets. *Acta Arith.*, 60(2):191–202, 1991.

[Wol09] J. Wolf. The structure of popular difference sets. *Israel J. Math.*, 2009. To appear.

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, England

E-mail address: `t.sanders@dpmms.cam.ac.uk`